2

Measurement and Uncertainty

2–1 BASIC NATURE OF MEASURING PROCESS

Measurement is the process of quantifying our experience of the external world. The nineteenth-century Scottish scientist, Lord Kelvin, once said that "when you can measure what you are speaking about and express it in numbers, you know something about it; but, when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the stage of science." While this may be a slight overstatement, it remains true that measurements constitute one of the basic ingredients of experimenting. We shall not reach a satisfactory level of competence in experimenting without knowledge of the nature of measurement and the significance of measurement statements.

It is obvious that the quantifying process will almost invariably involve comparison with some reference quantity (how many paces wide is my back yard?). It is equally obvious that the good order of society requires extensive agreement about the choice of reference quantities. The question of such measurement standards, defined by legislation and subject to international agreement, is extensive and important. No one seriously interested in measurement can ignore the question of defining and realizing standards in his area of work. A discussion of this important topic here would, however, distract us from our chief concern, the process of measuring. We shall, therefore, leave the topic of

standards without further mention except reference to the texts listed in the Bibliography, and take up the study of actual measuring processes.

Let us start at the most basic level with an apparently simple measurement; let us try to find out what kind of process is involved and what kind of statement can be made. If I give the notebook in which this is being written to someone and ask him to measure its length with a meter stick, the answer is absolutely invariable—the length of the notebook is 29.5 cm. But that answer must make us wonder: are we really being asked to believe that the length of the book is exactly 29.50000000 . . . . . cm? Surely not; such a claim is clearly beyond the bounds of credibility. So how are we to interpret the answer? A moment's thought in the presence of the notebook and a meter stick will make us realize that, far from determining the "right" or "exact" value, the only thing we can realistically do is approach the edge of the notebook along the scale, saying to ourselves as we go: "Am I sure the answer lies below 30 cm? Below 29.9 cm? Below 29.8 cm?" The answer to each of these questions will undoubtedly be "Yes." As we progress along the scale, however, we shall eventually reach a point at which we can no longer give the same confident reply. At that point we must stop, and we identify thereby one end of an interval that will become our measured value. In a similar way we can approach the edge of the notebook from below, asking ourselves at each stage: "Am I sure that the answer lies above 29.0 cm? 29.1 cm," and so on. Once again we shall reach a value at which we must stop, because we can no longer say with confidence that the answer lies above it. By the combination of these two processes we identify an interval along the scale. It is the smallest interval that, as far as we can be certain, does contain our desired value; within the interval, however, we do not know where our answer lies. Such is the only realistic outcome of a measuring process. We cannot look for exact answers, and we must be content with measured values that take the form of intervals. Not only does this example illustrate the essential nature of a measuring process, it also provides guidance for actually making measurements. The process of approaching the value we seek from each side separately reminds us of the necessity of stating the result as an interval, and also makes it easier to identify the edges of that interval.

The final outcome of our discussion is a most important one. As we make measurements and as we report the results we must keep in mind constantly this fundamental and vital point—measurements are not exact, single numbers but consist of intervals, within which we are confident that our desired value lies. The act of measurement requires us to determine both the location and width of this interval, and we do it by the careful exercise of visual judgment every time we make a measurement. There are no rules for determining the size of the interval, for it will depend on many factors in the measuring proc-

ess. The type of measurement, the fineness of the scale, our visual acuity, the lighting conditions—all will play a part in determining the width of the measurement interval. The width, therefore, must be determined explicitly each time a measurement is made. For example, it is a common error to believe that, when making a measurement using a divided scale, the "reading error" is automatically one half of the finest scale division. This is an erroneous oversimplification of the situation. A finely divided scale used to measure an object with ill-defined edges can give a measurement interval as large as several of the finest scale divisions; a well-defined object and good viewing conditions, on the other hand, may permit the identification of a measurement interval well within the finest scale division. Every situation must be assessed individually.

2–2 DIGITAL DISPLAY AND ROUNDING OFF

Other aspects may also confuse the issue. Consider, for example, a piece of equipment which gives a digital readout. If a digital voltmeter tells us that a certain potential difference is 15.4 V, does it intend to imply that the value is 15.40000 . . . exactly? Clearly not, but what does it mean? That depends on circumstances. If the instrument is made in such a way that it reads 15.4 V because the actual value is closer to 15.4 than it is to 15.3 or 15.5, then the meaning is: this reading lies between 15.35 and 15.45. On the other hand, a digital clock may be made in such a way that it changes its indication from 09.00 to 09.01 at the time of 09.01. When we see it reading 09.00, then, we know that the time lies between 09.00 and 09.01, a slightly different interpretation from that appropriate to the digital voltmeter. Again, each situation must be judged by itself.

These two examples of digital display illustrate a more general concept, the inaccuracy inherent in the process of "rounding off." Even without inaccuracy arising from limited ability to make measurements, a mere statement of a numerical quantity can contain inaccuracy. Consider the statement

$$\pi = 3.14$$

We all know that this is not so because we can remember some, at least, of the following numbers, 3.14159 . . . . So what can we mean by quoting π as 3.14? It can mean only that π has a value closer to 3.14 than it does to 3.13 or 3.15. Our statement is, therefore, that π lies between 3.135 and 3.145. This range of possibility represents what is sometimes known as a "rounding-off error." Such errors can be small and unimportant, or they can become significant. In a long calculation, for example, there is a chance that rounding-off errors can accumulate, and it becomes wise, especially in these days of conveniently available calculators, to carry through the calculation more

figures than one might think would be necessary. A similar rounding-off error can appear in statements about measurement. We sometimes hear that someone has made a measurement on a scale which was "read to the nearest millimeter" or some such phrase. This is not a very good way of reporting a measurement because it obscures the actual value of the measurement interval. We do, however, encounter such statements and, if we are obliged to deal with a measurement quoted in that form, we can only assume that the scale division quoted represents some kind of minimum value for the size of the measurement interval.

2–3 ABSOLUTE AND RELATIVE UNCERTAINTY

By whatever means we have made a measurement, the final outcome should be an interval which represents, to the best of our ability, the range inside which the desired value lies. In the example we used first the experimenter might be able to state with confidence no more than that the length of the notebook lay between 29.4 and 29.6 cm. Although the only meaningful outcome of a measuring process consists of such an interval or range, it is frequently desirable, for purposes of description or further calculation, to rephrase the quoted value. We take the interval 29.4–29.6 and *rename* it 29.5 ± 0.1 cm. Although obviously no more than a renamed expression of the original interval, the new form does offer certain advantages. It gives us a central value, 29.5, which can be used in further calculations. It also gives us a value, ± 0.1, called the "uncertainty" of the measurement, by which we can judge the quality of the measuring process and which can be used in separate calculations on uncertainties. One disadvantage in this mode of expression is the return to a central value, 29.5. Unless we remember clearly that only the complete quantity, 29.5 ± 0.1, serves as an adequate statement of the answer, we may become sloppy in making and reporting measurements and may forget the essential presence of the uncertainty. We should all make it an invariable practice to associate an uncertainty value with a reading, both at the time we make the measurement, and subsequently, whenever the value is quoted or used in further calculation.

Since the figure ± 0.1 cm represents the actual amount, or range, by which the reading of 29.5 is uncertain, it is often called the "absolute uncertainty" of the reading, and we shall consistently use this terminology. In addition, other aspects soon become important. How significant is an uncertainty of ± 0.1 cm? When we are measuring the length of a notebook, it is significant to a certain extent. If we are measuring the distance between two cities, an uncertainty of ± 0.1 cm is probably completely insignificant. If, on the other hand, we are measuring the size of a microscopic bacterium, an uncertainty of

± 0.1 cm would make the measurement meaningless. For this reason, it is frequently desirable to compare an uncertainty figure with the actual value of the measurement; by so doing the significance of the uncertainty can be realistically assessed. We define the ratio

$$\textbf{relative uncertainty} = \frac{\textbf{absolute uncertainty}}{\textbf{measured value}}$$

In the case of our example

$$\text{relative uncertainty} = \pm \frac{0.1}{29.5} = \pm 0.003$$

This relative uncertainty is often quoted as a percentage, so that, in the present case, the relative uncertainty would be $\pm 0.3\%$. Such a quantity gives us a much better feeling for the quality of the reading, and we often call it the "precision" of the measurement. Note that the absolute uncertainty has the same dimensions and units as the basic measurement (29.5 cm is uncertain by 0.1 cm), while the relative uncertainty, being a ratio, has no dimensions or units and is a pure number.

2–4 SYSTEMATIC ERROR

The kind of uncertainty that we have been considering arises from naturally occurring inadequacy in the measuring process. A different type of error can appear when something affects all the measurements of a series in an equal or a consistent way. For example, a voltmeter or a micrometer caliper can have a zero error, a wooden meter stick may have shrunk, a person may consistently press a stopwatch button $\frac{1}{10}$ sec behind the event, and so on. These errors are termed "systematic errors," a subclass of which are "calibration errors." Because such systematic errors may not be immediately visible as one makes a measurement, it is necessary to be vigilant and remember at all times the possibility of their presence. Instrumental zeroes, for example, should automatically be checked every time an instrument is used. Although it may be less easy to check calibration, the accuracy of electrical meters, stopwatches, thermometers, and other such instruments should not be taken for granted and should be checked whenever possible. Also, the presence on an instrument of a precise-looking, digital readout with four or five supposedly significant figures should not be taken as proof of precision and freedom from systematic error. Most of a batch of electronic timers that our laboratory recently acquired for laboratory teaching, which could supposedly measure time intervals with millisecond precision, turned out to have calibration errors as large as 14%. Do not be deceived; view all measuring instruments with suspicion and check instrumental calibration whenever possible.

2–5 UNCERTAINTY IN CALCULATED QUANTITIES

In the preceding sections we have been concerned solely with the concept of uncertainty in a single measurement. It is rare, however, that a single measurement ends the process. Almost invariably the result we desire is a combination of two or more measured quantities or is, at least, a calculated function of a single measurement. We might wish, for example, to calculate the cross-sectional area of a cylinder from a measurement of its diameter, or its volume from measurements of both diameter and length. The various measurements will sometimes be of different types, as in a calculation of g from values of the length and period of a pendulum. In these cases the presence of uncertainty in the basic measurements will obviously entail the presence of uncertainty in the final computed value. It is this final uncertainty that we now wish to calculate. For the purposes of this section we shall assume that our uncertainties have the character of ranges or intervals within which we are "almost certain" that our answer lies. For the computed values we shall calculate intervals within which we wish, once again, to be "almost certain" that our answer lies. That means that we must do our calculation for the "worst case" of combined uncertainties. This is perhaps a pessimistic assumption, and we shall see later, in Chap. 3, how the probabilities associated with various error combinations enable us to make a more realistic and less pessimistic estimate. For the moment, however, let us assume that we wish to calculate, from the uncertainties in the primary values, the maximum range of possibility for the computed answer.

2–6 UNCERTAINTY IN FUNCTIONS OF ONE VARIABLE ONLY

Consider a measured quantity x_0 with an uncertainty $\pm\delta x$, and consider a computed result z to be a function of the variable x. Let

$$z = f(x)$$

This function enables us to calculate the required value z_0 from a measured value x_0. Moreover, the possibility that x can range from $x_0 - \delta x$ to $x_0 + \delta x$ implies a range of possible values of z from $z_0 - \delta z$ to $z_0 + \delta z$. We now wish to calculate the value of δz. The situation is illustrated graphically in Fig. 2–1, in which, for a given $f(x)$, we can see how the measured value x_0 gives rise to the computed result z_0, and how the range $\pm\delta x$ about x_0 produces a corresponding range $\pm\delta z$ about z_0.

Before considering general methods of evaluating δz it is instructive to see how finite perturbations are propagated in simple functions. Consider, for example, the function

$$z = x^2$$

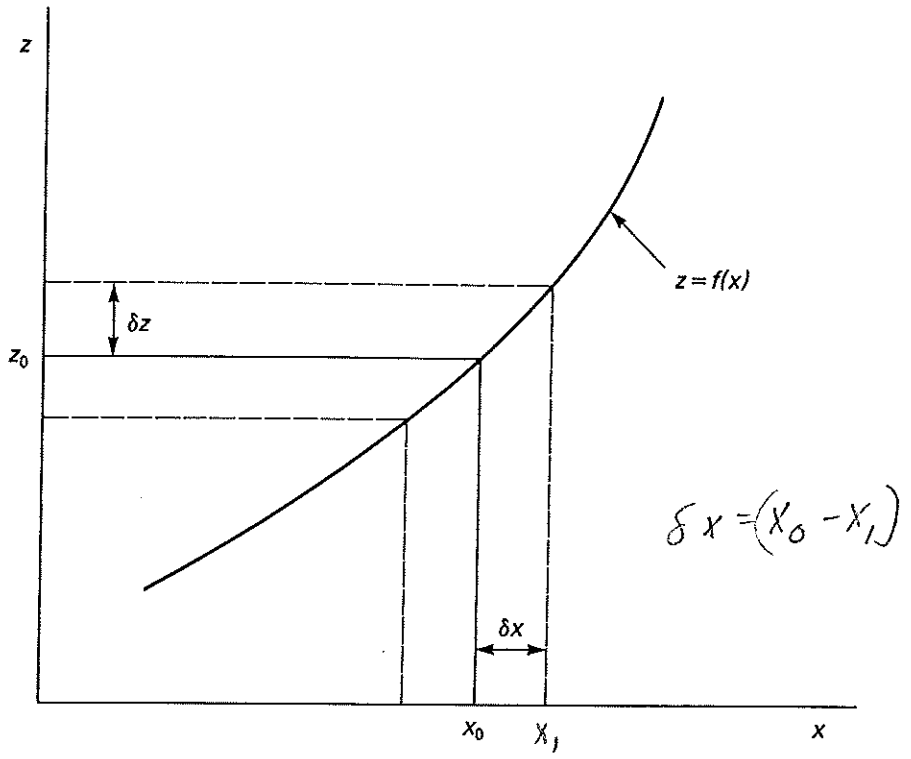

Figure 2–1 Propagation of uncertainty from one variable to another.

If x can range between $x_0 - \delta x$ and $x_0 + \delta x$, then z can range between $z_0 - \delta z$ and $z_0 + \delta z$, where

$$\begin{aligned} z_0 \pm \delta z &= (x_0 \pm \delta x)^2 \\ &= x_0^2 \pm 2x_0\delta x + (\delta x)^2 \end{aligned}$$

We can ignore $(\delta x)^2$, since δx is assumed to be small compared with x_0, and equate z_0 to x_0^2, giving for the value of δz

$$\delta z = 2x_0\delta x$$

This can more conveniently be expressed in terms of the relative uncertainty $\delta z/z_0$:

$$\frac{\delta z}{z_0} = \frac{2x_0\delta x}{x_0^2} = 2\frac{\delta x}{x_0}$$

Thus, the relative uncertainty of the computed result is twice that of the initial measurement.

Although it is essential to bear in mind the nature of propagated uncertainty, as illustrated by the use of finite differences, considerable simplification of the formulation can be achieved using differential calculus.

2–7 GENERAL METHOD FOR UNCERTAINTY IN FUNCTIONS OF A SINGLE VARIABLE

In the preceding section the finite differences δz and δx are merely an expression of the derivative dz/dx. We can therefore obtain our value of δz by first using standard techniques to obtain dz/dx in the form

$$\frac{dz}{dx} = \frac{d(f(x))}{dx}$$

and then writing

$$\delta z = \frac{d(f(x))}{dx}\,\delta x \tag{2–1}$$

This is a relatively simple procedure, and it will work well in cases for which the elementary, finite-difference approach would lead to algebraic complexity. Thus, if

$$z = \frac{x}{(x^2 + 1)}$$

then

$$\frac{dz}{dx} = \frac{x^2 + 1 - x \cdot 2x}{(x^2 + 1)^2}$$

$$= \frac{1 - x^2}{(1 + x^2)^2}$$

and

$$\delta z = \frac{1 - x^2}{(1 + x^2)^2}\,\delta x$$

This calculation would have been very awkward by any other approach. Furthermore, it gives δz generally as a function of x and δx; any particular desired value can be obtained by setting $x = x_0$. Let us now use this technique to evaluate uncertainties for some common functions.

(a) Powers

Consider

$$z = x^n$$

$$\frac{dz}{dx} = nx^{n-1}$$

$$\delta z = nx^{n-1}\delta x$$

The significance of this result becomes a little more obvious when expressed in terms of the relative uncertainty. Thus,

$$\frac{\delta z}{z} = n\frac{\delta x}{x}$$

Thus, when evaluating powers, the *relative uncertainty* of the result is the relative uncertainty of the basic quantity multiplied by the power involved. This will be valid for either powers or roots, so that precision diminishes as a quantity is raised to powers and improves on taking roots. This situation must be carefully watched in an experiment in which powers are involved. The higher the power, the greater will be the need for high initial precision.

(b) Trigonometric Functions

We shall do only one example, since all the others can be treated in similar ways. Consider

$$z = \sin x$$

Here

$$\frac{dz}{dx} = \cos x$$

and

$$\delta z = (\cos x)\ \delta x$$

This is one case where the elementary method of inserting $x_0 \pm \delta x$ shows the result more clearly. Using the approximation

$$\cos \delta x = 1$$

we obtain

$$\delta x = \cos x \sin \delta x$$

showing that the δx in the preceding result is really $\sin \delta x$ in the limit of small angles. Only in the case of very large uncertainty would this difference be significant, but it is best to understand the nature of the result. Clearly δx should be expressed in radian measure. The result will normally have straightforward application when dealing with apparatus such as spectrometers.

(c) Logarithmic and Exponential Functions

Consider

$$z = \log x$$

Here

$$\frac{dz}{dx} = \frac{1}{x}$$

and

$$\delta z = \frac{1}{x}\delta x$$

The relative uncertainty can be calculated as usual. If

$$z = e^x$$

$$\frac{dz}{dx} = e^x .$$

and

$$\delta z = e^x \delta x$$

This is an important case, since exponential functions occur frequently in science and engineering. These functions can become very sensitive to the exponent when it takes values much over unity, and the uncertainty δz may become very large. This will be familiar, for example, to anyone who has watched the current fluctuations in a thermionic diode that can result from quite small variations in filament temperature.

As stated earlier, the method can be easily applied to any function not listed above by evaluating the appropriate derivative and using Eq. (2–1).

2–8 UNCERTAINTY IN FUNCTIONS OF TWO OR MORE VARIABLES

If the result is to be computed from two or more measured quantities, x, y, etc., the uncertainty in the result can, as was mentioned in Sec. 2–5, be regarded in two different ways. We could be as pessimistic as possible and suppose that the actual deviations of x and y happen to combine in such a way as to drive the value of z as far as possible from the central value. We would, in this way, calculate a value for δz which gives the extreme width of the range of possible z values. On the other hand we can argue that it is more probable for the uncertainties in the basic measurements to combine in a less extreme way, some making positive contributions to δz and some negative, so that the resulting δz value is smaller than for the pessimistic assumption. This argument is valid, and we shall deal later with the question of probable uncertainty in computed quantities. For the moment, however, let us calculate that value

of δz which represents the widest range of possibility for z. Such an approach, if pessimistic, is certainly safe, since, if δx, δy, etc., represent limits within which we are "almost certain" the actual values lie, then the calculated δz will give those limits within which we are equally certain that the actual value of z lies.

The most instructive initial approach uses the elementary substitution method, and we shall use this for the first two functions.

(a) Sum of Two or More Variables

Consider

$$z = x + y$$

The uncertainty in z will be obtained from

$$z_0 \pm \delta z = (x_0 \pm \delta x) + (y_0 \pm \delta y)$$

and the maximum value of δz is obtained by choosing similar signs throughout. Thus,

$$\delta z = \delta x + \delta y$$

As might be expected, the uncertainty in the sum is just the sum of the individual uncertainties. This can be expressed in terms of the relative uncertainty,

$$\frac{\delta z}{z} = \frac{\delta x + \delta y}{x + y}$$

but no increased clarification is achieved.

(b) Difference of Two Variables

Consider

$$z = x - y$$

As in the case above, δz will be obtained from

$$z_0 \pm \delta z = (x_0 \pm \delta x) - (y_0 \pm \delta y)$$

Here, however, we can obtain the maximum value of δz by choosing the *negative* sign for δy, giving, once again,

$$\delta z = \delta x + \delta y$$

We can see from this equation that, when x_0 and y_0 are close together and $x - y$ is small, the relative uncertainty can rise to very large values. This is, at best, an unsatisfactory situation, and the precision can be low enough to de-

stroy the value of the measurement. The condition is particularly hazardous since it can arise unnoticed. It is perfectly obvious that, if it were possible to avoid it, no one would attempt to measure the length of my notebook by measuring the distance of each edge from a point a mile away and then subtracting the two lengths. However, it can happen that a desired result is to be obtained by subtraction of two measurements made separately (two thermometers, clocks, etc.), and the character of the measurement as a difference may not be strikingly obvious. Consequently, all measurements involving differences should be treated with the greatest caution. The way to avoid the difficulty, clearly, is to measure the difference directly, rather than obtaining it by subtraction between two measured quantities. For example, if one has an apparatus within which two points are at potentials above ground of $V_1 = 1500$ V and $V_2 = 1510$ V, respectively, and the required quantity is $V_2 - V_1$, only a very high quality voltmeter would permit the values of V_2 and V_1 to be measured with the exactness required to achieve even 10% precision in $V_2 - V_1$. On the other hand, an ordinary 10-V table voltmeter, connected between the two points and measuring $V_2 - V_1$ directly, will immediately give the desired result with 2% or 3% precision.

2–9 GENERAL METHOD FOR UNCERTAINTY IN FUNCTIONS OF TWO OR MORE VARIABLES

The last two examples, treated by the elementary method, suggest that, once again, the differential calculus may offer considerable simplification of the treatment. It is clear that, if we have

$$z = f(x, y)$$

the appropriate quantity for calculating δz is the total differential dz. This is given by

$$dz = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy \tag{2–2}$$

We shall take this differential and treat it as a finite difference δz that can be calculated from the uncertainties δx and δy. Thus,

$$\delta z = \frac{\partial f}{\partial x} \delta x + \frac{\partial f}{\partial y} \delta y$$

and the derivatives $\partial f/\partial x$ and $\partial f/\partial y$ will normally be evaluated for the values, x_0 and y_0, at which δz is required. We may find that, depending on the function f, the sign of $\partial f/\partial x$ or $\partial f/\partial y$ turns out to be negative. In this case, using our pessimistic requirement for the maximum value of δz, we should choose nega-

tive values for the appropriate δx or δy, obtaining thereby a wholly positive contribution to the sum.

(a) Product of Two or More Variables

Suppose

$$z = xy$$

To use Eq. (2–2) we need the values of $\partial z/\partial x$ and $\partial z/\partial y$. They are

$$\frac{\partial z}{\partial x} = y \qquad \text{and} \qquad \frac{\partial z}{\partial y} = x$$

Thus, the value of δz is given by

$$\delta z = y\,\delta x + x\,\delta y$$

The significance of this result is more clearly seen when it is converted to the relative uncertainty

$$\frac{\delta z}{z} = \frac{\delta x}{x} + \frac{\delta y}{y}$$

Thus, when the desired quantity is a product of two variables, its relative uncertainty is the sum of the relative uncertainties of the components.

The most general case of a compound function, very commonly found in physics, involves an algebraic product that has components raised to powers. Let

$$z = x^a y^b$$

where a and b may be positive or negative, integral or fractional. In this case the formulation is greatly simplified by taking logs of both sides before differentiating. Thus,

$$\log z = a \log x + b \log y$$

whence, differentiating implicitly,

$$\frac{dz}{z} = a\frac{dx}{x} + b\frac{dy}{y}$$

As usual, we take the differentials to be finite differences and obtain

$$\frac{\delta z}{z} = a\frac{\delta x}{x} + b\frac{\delta y}{y}$$

Note that this process gives the relative uncertainty directly, and this is fre-

quently convenient. If the absolute uncertainty δz is required, it can be evaluated simply by multiplying the relative uncertainty by the computed value z_0, which is normally available. This form of implicit differentiation still offers the simplest procedure even when z itself is raised to some power. For, if the equation reads

$$z^2 = xy$$

it is unnecessary to rewrite it

$$z = x^{1/2}y^{1/2}$$

and work from there, because, by taking logs,

$$2 \log z = \log x + \log y$$

whence

$$2\frac{\delta z}{z} = \frac{\delta x}{x} + \frac{\delta y}{y}$$

giving $\delta z/z$ as required.

(b) Quotients

These can be treated as products in which some of the powers are negative. As before, the maximum value of δz will be obtained by neglecting negative signs in the differential and combining all the terms additively.

If a function other than those already listed is encountered, some kind of differentiation will usually work. It is frequently convenient to differentiate an equation implicitly, thereby avoiding the requirement to calculate the unknown quantity explicitly as a function of the other variables. For example, consider the thin-lens equation

$$\frac{1}{f} = \frac{1}{o} + \frac{1}{i}$$

where the focal length f is a function of object distance o and image distance i, the measured quantities. We can differentiate the equation implicitly to obtain

$$-\frac{df}{f^2} = -\frac{do}{o^2} - \frac{di}{i^2}$$

It is now possible to calculate df/f directly and more easily than by writing f explicitly as a function of o and i and differentiating. In this way we can prepare a formula for the uncertainty into which all the unknowns can be inserted directly. Make sure that appropriate signs are used so that all contributions to the uncertainty add positively to give outer limits of possibility for the answer.

If the function is so big and complicated that we cannot obtain a value for δz in general, we can always take the measured values, x_0, y_0, etc., and work out z_0. We can then work out two different answers, one using the actual, numerical values of $x_0 + \delta x$, $y_0 + \delta y$ (or $y_0 - \delta y$ if appropriate), etc., to give one of the outer values of z and the other using $x_0 - \delta x$, etc. These two values will correspond to the limits on z, and we shall know the value of δz.

2–10 COMPENSATING ERRORS

A special situation can appear when compound variables are involved. Consider, for example, the well-known relation for the angle of minimum deviation D_m for a prism of refractive index n and vertical angle A:

$$n = \frac{\sin \frac{1}{2}(A + D_m)}{\sin \frac{1}{2}A}$$

If A and D_m are measured variables with uncertainties δA and δD_m, the quantity n will be the required answer, with an uncertainty δn. It would be fallacious, however, to calculate the uncertainty in $A + D_m$, then in $\sin \frac{1}{2}(A + D_m)$, and combine that with the uncertainty in $\sin \frac{1}{2}A$, treating the function as a quotient of two variables. This can be seen by thinking of the effect on n of an increase in A. *Both* $\sin \frac{1}{2}(A + D_m)$ and $\sin \frac{1}{2}A$ increase, and the change in n is not correspondingly large. The fallacy lies in applying the methods of the preceding sections to variables that are not independent (e.g., $A + D_m$ and A). The cure is either to reduce the equation to a form in which the variables are all independent, or else to go back to first principles and use Eq. (2–2) directly. Cases which involve compensating errors should be watched carefully, since they can, if treated incorrectly, give rise to errors in uncertainty calculations that are hard to detect.

2–11 SIGNIFICANT FIGURES

Since computations tend to produce answers consisting of long strings of numbers, we must be careful to quote the final answer sensibly. If, for example, we are given the voltage across a resistor as 15.4 ± 0.1 volts and the current as 1.7 ± 0.1 amps, we can calculate a value for the resistance. The ratio V/I comes out on my calculator as 9.0588235 ohms. Is this the answer? Clearly not. A brief calculation shows that the absolute uncertainty in the resistance is close to 0.59 ohms. So, if the first two places of decimals in the value for the resistance are uncertain, the rest are clearly meaningless. A statement like $R = 9.0588235 \pm 0.59$ ohms is, therefore, nonsense. We should quote our results in such a way that the answer and its uncertainty are consistent, e.g., $R = 9.06 \pm 0.59$ ohms.

But is even this statement really valid? Remember that the originally quoted uncertainties for V and I had the value ± 0.1, containing one significant figure. If we did now know these uncertainties any more precisely, we have no right to claim two significant figures for the uncertainty in R. Our final, valid, and self-consistent statement is, therefore, R = 9.1 ± 0.6 ohms. Only if we had real reason to believe that our original uncertainty was accurate to two significant figures could we lay claim to two significant figures in the final uncertainty and a correspondingly more precisely quoted value for R. In general terms we must make sure that our quoted values for uncertainty are consistent with the precision of the basic uncertainties and that the number of quoted figures in the final answer is consistent with the uncertainty of that final answer. We must avoid statements like $z = 1.234567 \pm 0.1$ or $z = 1.2 \pm 0.000001$.

PROBLEMS

1. I use my meter stick to measure the length of my desk. I am sure that the length is not less than 142.3 cm and not more than 142.6 cm. State this measurement as a central value ± uncertainty. What is the relative uncertainty of the measurement?
2. I read a needle-and-scale voltmeter and ammeter and assess the range of uncertainty visually. I am sure the ammeter reading lies between 1.24 and 1.25 A and the voltmeter reading between 3.2 and 3.4 V. Express each reading as a central value ± uncertainty and evaluate the relative uncertainty of each measurement.
3. My digital watch gives a time reading as 09:46. What is the absolute uncertainty of the measurement?
4. If I can read a meter stick with absolute uncertainty ± 1 mm, what is the shortest distance that I can measure if the relative uncertainty is not to exceed (a) 1%, (b) 5%?
5. I use a thermometer graduated in $\frac{1}{5}$ degree Celsius to measure outside air temperature. Measured to the nearest $\frac{1}{5}$ degree, yesterday's temperature was 22.4° Celsius and today's is 24.8° Celsius. What is the relative uncertainty in the temperature difference between yesterday and today?
6. The clock in the lab has a seconds hand that moves in one-second steps. I use it to measure a certain time interval. At the beginning of the interval it reads 09:15:22 (hours:minutes:seconds) and at the end it reads 09:18:16. What is the relative uncertainty of the measured time interval?
7. For the desk mentioned in Problem 1 I measure the width, and I am sure the measurement lies between 78.2 cm and 78.4 cm. What is the absolute uncertainty of the calculated area of the desk top?
8. In measuring the resistance of a resistor, the voltmeter reading was 15.2 ± 0.2 V and the ammeter reading was 2.6 ± 0.1 A. What is the absolute uncertainty of the resistance calculated using the equation $R = V/I$?

9. A simple pendulum is used to measure the acceleration of gravity using $T = 2\pi\sqrt{l/g}$. The period T was measured to be 1.24 ± 0.02 sec and the length to be 0.381 ± 0.002 m. What is the resulting value for g with its absolute and relative uncertainty?

10. An experiment to measure the density, d, of a cylindrical object uses the equation $d = m/\pi r^2 l$, where

$$m = \text{mass} = 0.029 \pm 0.005 \text{ kg}$$

$$r = \text{radius} = 8.2 \pm 0.1 \text{ mm}$$

$$l = \text{length} = 15.4 \pm 0.1 \text{ mm}$$

What is the absolute uncertainty of the calculated value of the density?

11. The focal length, f, of a thin lens is to be measured using the equation $1/o + 1/i = 1/f$, where

$$o = \text{object distance} = 0.154 \pm 0.002 \text{ m}$$

$$i = \text{image distance} = 0.382 \pm 0.002 \text{ m}$$

What is the calculated value for focal length, its absolute uncertainty, and its relative uncertainty?

12. A diffraction grating is used to measure the wavelength of light using the equation $d \sin \theta = \lambda$. The value of θ is measured to be $13° 34' \pm 2'$. Assuming that the value of d is 1420×10^{-9} m and that its uncertainty can be ignored, what are the absolute and relative uncertainties in the value of λ?

13. A value is quoted as 14.253 ± 0.1. Rewrite it with the appropriate number of significant figures. If the value is quoted as 14.253 ± 0.15, how should it be written?

14. A value is quoted as $6.74914 \pm 0.5\%$. State it as a value $\pm$ absolute uncertainty, both with the appropriate number of significant figures.

3

Statistics of Observation

3–1 STATISTICAL UNCERTAINTY

In the preceding chapter we considered measurements in which the uncertainty could be estimated by personal judgment. In these, supposing that we have judged the situation accurately, repeated measurements should give consistent answers. Sometimes, however, repeated measurements give clearly different answers. For example, if we are using a Geiger counter and scaler to measure the activity of a radioactive source, and we decide, with given geometry, to obtain the number of counts in a 10-second interval, we would find that the results obtained by counting in successive 10-second intervals are *not* the same. We can encounter the same situation in measurements that involve visual judgment. If, for example, we wish to find the image formed by a thin lens, we may be unable to judge the position of the image accurately enough to obtain repeatedly the same reading on a good, finely divided distance scale. Whether the fluctuation is intrinsic to the system under investigation (as in the radioactive source, where the fluctuation arises from the basic nature of radioactive decay) or arises from our difficulty in making a measurement, we must find out how to make sensible statements about measurements that show such fluctuations.

What kind of statement will it be possible to make? No longer can we make such statements as we made earlier having the form "I am virtually certain that the answer lies within the interval . . . ." In fact, apart altogether

from the impossibility of obtaining "right" answers, we shall find out that the difficulty lies not so much in constructing sensible answers as in knowing the sensible questions to ask. We shall discover that the only sensible questions involve, as before, intervals along our scale of values—this time, however, interpreted in terms of probabilities instead of certainty. Our search for a solution will be fairly lengthy, but at the end the answer will turn out to be simple and elegant.

To start our search, let us go back to the basic situation. Let us assume that we have made a single measurement and, in order to check our work, that we have made the measurement a second time and obtained a different answer. What are we supposed to do? We have no way of saying that one answer is "right" and the other "wrong." Which one would we choose to be "right"? In response to this ambiguity the natural reaction would be to try a third time, hoping, perhaps, that the third reading will confirm one or other of the first two. Very likely it will not be so obliging and will simply add to the confusion by supplying a third possibility. Faced with growing complexity, we could decide to keep on making measurements to see what happens. Let us suppose that our curiosity has prompted us to make a substantial number of repeated measurements, say 100, and we now ask: what is the answer? As was mentioned earlier, it is more significant to ask: what is the question? That depends very much on the use to which we wish to put the measurements. A physicist measuring the position of an optical image may be seeking something he would like to consider as the "right" answer. A person measuring the activity of a radioactive source may wish to use it in a way that requires him to know the number of counts he will obtain in a certain 10-second interval tomorrow. A sociologist counting political opinions wishes to predict the outcome of the next election, etc. There is no single question and no unique answer. The treatment we give our fluctuating numbers depends on circumstances. Let us now consider some of the possibilities.

3–2 HISTOGRAMS AND DISTRIBUTIONS

Let us assume that we have made 100 measurements of some quantity and that we must now report our results. The first response to the question, "What did you obtain?" is the rather feeble reply, "I made the measurement 100 times and here are the 100 answers." This is perhaps free of error but is hardly helpful. Our audience will find it difficult to make any sense out of a plain list of numbers, and questions will naturally arise, such as: are there any regularities in the numbers, do any appear more frequently than others, etc.? In order to show the characteristics of the measurements more clearly, some kind of graphic display would clearly be helpful.

One common mode of presentation is the histogram. To construct this diagram we divide the scale along which the measurements are spread into intervals, and we count the readings that fall within each interval. We then plot these numbers on a vertical scale against the intervals themselves. It is conventional to use a bar diagram to indicate the number of readings, and the result will be similar to Fig. 3–1. At once we improve our comprehension of the measurements enormously, because we can see at a glance how the values are *distributed* along the scale. This distribution is the key to satisfactory interpretation of the measurements. Usually we find that the readings tend to occur more frequently in the middle of the range, and, if this is so and we are unable to make any other sensible statement, we can always content ourselves with the simple assertion that the observations have "central tendency." This may suffice, and when we have drawn the histogram we may be able to stop. Many

Table 1

85	109	114	121	127	131
92	109	114	121	127	132
96	110	114	122	127	133
97	110	115	122	127	134
97	111	116	122	128	134
97	111	116	122	128	134
100	111	116	122	128	134
101	111	117	123	128	135
101	111	117	123	128	136
102	112	118	123	128	137
102	112	118	123	130	137
103	112	119	123	130	137
103	113	119	124	130	144
105	113	120	124	130	148
106	113	120	124	130	149
106	113	120	125	130	
107	113	120	125	131	
108	113	121	125	131	
108	114	121	126	131	

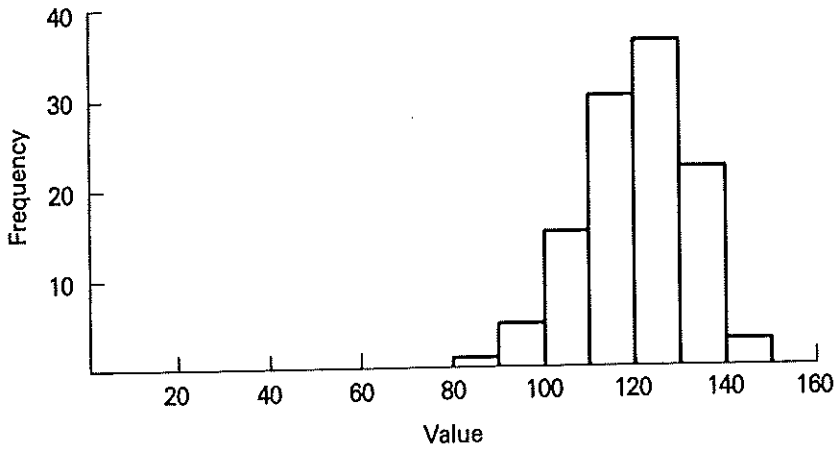

Figure 3-1 A set of observations and its histogram.

results from measurement processes are presented simply by offering the histogram; the reader can view the distribution and draw his own conclusions.

3–3 CENTRAL VALUES OF DISTRIBUTIONS

Frequently, however, we wish to go further and, as a substitute for the whole histogram, find some shorthand way of describing the distribution without actually showing the whole diagram. We can seek answers, therefore, to questions such as: what *single* number best characterizes the complete group of observations? There are several candidates for such designation, and we choose one on the basis of the future use of the information. The various possibilities are:

(a) Mode

Most distributions have a peak near the center. If this peak is well defined, the value on the horizontal scale at which it occurs is called the **mode** of the distribution. Whenever we wish to draw attention to such central concentration in our measured values, we quote the modal value. Sometimes a distribution will show two peaks; we call it a **bimodal** distribution and quote the two modal values.

(b) Median

If we place all our readings in numerical order and divide them in the middle into two equal parts, the value at which the dividing line comes is called the **median.** Since it is obvious that areas under distribution graphs represent numbers of observations (the left-hand bar in Fig. 3–1 represents 5 observations, the second from the left represents 9, so that the two together represent 14, and so on), the median is that value at which a vertical line divides the distribution into two parts of equal area. The median is frequently quoted in sociological work; people talk about median salaries for certain groups of employees, etc.

(c) Mean

The third of the commonly quoted numbers is the familiar arithmetic average or mean. For a group of N observations, x_i, the mean $\bar{x}$ is defined by

$$\bar{x} = \frac{\Sigma x_i}{N} \tag{3–1}$$

We shall discover that, for our purposes, the mean is the most useful of the three quantities we have defined.

Notice that, for a symmetrical distribution, the mean, median, and mode all coincide at the center of the distribution. If, on the other hand, the distribution is not symmetrical, each will have a separate value. For the histogram shown in Fig. 3–1, the values of the mean, median, and mode are shown in Fig. 3–2, which illustrates their relationship to the distribution. If the distribution is markedly asymmetric, the difference between the mode, median, and mean can be substantial. Consider, for example, the distribution of family income in a country. The presence of the millionaires, although few in number, has an effect on the mean that counterbalances many members of the popula-

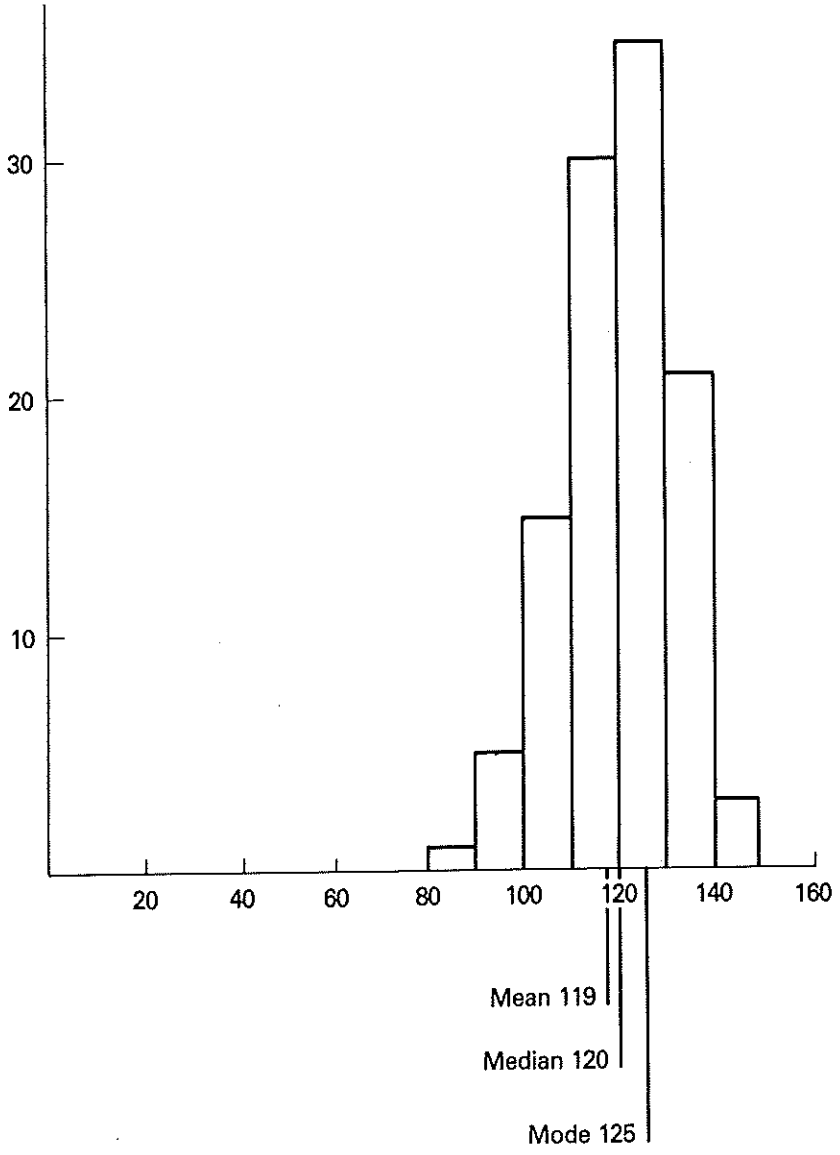

Figure 3-2 The relationship between a histogram and its mean, median, and mode.

tion at the low end of the salary scale. The mode and the mean thus differ substantially. This example illustrates the care required in interpreting quoted statistics; people who quote statistics frequently do so in the way that best suits their particular purpose.

3–4 THE BREADTH OF DISTRIBUTIONS

Let us now turn to the question: to what extent is our chosen number representative of the distribution as a whole? That is, how reliable is it to use a single number to substitute for a whole distribution? At the present stage we have no justification to offer for the procedures that will be described. We shall rely, instead, on an intuitive feeling that, the broader is the distribution, the less significance we can ascribe to any one of the three central values. On the other hand, the narrower the distribution, the more we feel entitled to confidence in the mean, mode, or median as significant quantities for the distribution.

Let us, therefore, construct a quantity that will be a measure of the breadth of the distribution. We could invent many such quantities, but, for reasons that need not concern us at the moment, we shall define a quantity that is almost universally used. We define the **standard deviation** of the distribution, S, to be

$$S = \sqrt{\frac{\Sigma\,(\bar{x} - x_i)^2}{N}} \tag{3–2}$$

The definition is to some extent arbitrary, for, in defining a measure of the breadth of the distribution, we could have chosen other powers for the quantity $(\bar{x} - x_i)$, and we could have chosen other denominators. There are, however, reasons for these choices; these reasons and the significance of the standard deviation will become clear shortly.

We can pause at this stage to summarize the progress so far. If we have made repeated measurements of a quantity and wish to state the result in numerical terms, we can do a number of things: (a) we can show the histogram, (b) we can quote the mode, median, or mean as a measure of the location of the distribution, and (c) we can quote the standard deviation as a measure of the confidence we can place in the results. We sometimes leave the outcome of a measuring process in this form; the quantities involved are universally understood, and the procedure is acceptable.

For our present purpose, however, we seek more detailed, numerical interpretation of the quoted values.

3–5 SIGNIFICANCE OF THE MEAN AND STANDARD DEVIATION

In this and the following sections we shall, for reasons that will become clear, ignore the mode and median and restrict ourselves to numerical interpretation of the mean and the standard deviation. Since the presence of random fluctuation has denied us the opportunity to identify a realistic interval within which we can feel certain our answer lies, we must alter our expectations of the measuring process. As we have said before, it is not so much a matter of obtaining sensible answers to questions as of knowing the sensible questions to ask. Specifically, of course, it is not sensible to ask: what is the right answer? It is not even sensible to ask: having made one hundred observations of a quantity, what shall I obtain when I make the measurement the next time? The only sensible questions involve not certainty but probability, and several different questions about probabilities are possible.

We could ask, for example: what is the probability that the 101st reading will fall within a certain range on our scale of values? That is a sensible question, and sensible answers can easily be imagined. If, for example, of our 100 original readings, a certain fraction of the values fell within that particular range, we might feel justified in choosing that fraction as the probability we seek. This would not be an unrealistic guess, and we could attempt a standardized description of our distribution by quoting the fraction of the total number of readings that fall within a specified interval, such as $x \pm S$. This would satisfactorily convey information about our set of readings to other people, but a major problem appears when we discover that our answers for probabilities are specific to our particular histogram. If we were to make another series of 100 readings, holding all the conditions the same as they were before in the hope of obtaining the same histogram, we would be disappointed. The new histogram would not duplicate the first exactly. It might have similar general characteristics with respect to location and breadth, but its detailed structure would not be the same as before, and we would obtain different answers to questions about probabilities.

How, then, are we going to find answers to our questions that have some kind of widely understood numerical significance? One solution is to abandon the attempt to describe our particular histogram and to start talking about defined, theoretical distributions. These may have the disadvantage of uncertain relevance to our particular set of observations, but there is the enormous advantage that, since they are defined, theoretical constructs, they have properties that are definite, constant, and widely understood. Many such theoretical distributions have been constructed for special purposes, but we shall restrict ourselves to one only, the Gaussian or "normal" distribution.

We use the Gaussian distribution to interpret many kinds of physical measurement, partly because the mechanical circumstances of many physical measurements are in close correspondence with the theoretical foundations of the Gaussian distribution, and partly because experience has shown that Gaussian statistics do provide a reasonably accurate description of many real events. For only one common type of physical measurement is another distribution more appropriate; in counting events like radioactive decay we must use a distribution called the Poisson distribution, but, even for it, the difference from Gaussian statistics becomes significant only at low counting rates. Further information about Poisson statistics will be found in books describing experimental methods in nuclear or high-energy physics. Apart from these special cases, we can feel relatively confident that Gaussian statistics can be usefully applied to most real measurements. We should, however, always remember that, unless we actually test our measurements for correspondence with the Gaussian distribution, we are making an assumption that Gaussian statistics are applicable, and we should remain alert to any evidence that the assumption may be invalid.

3–6 GAUSSIAN DISTRIBUTION AND SAMPLING

Even if, to use it successfully, we need not know very much about the origins of the Gaussian distribution, it is interesting to know why its derivation makes it particularly relevant to many physical measurements. The Gaussian distribution can be derived from the assumption that the total deviation of a measured quantity, x, from a central value, X, is the consequence of a large number of small fluctuations that are of random occurrence. To construct a simple model of such a situation, let us suppose that there are m such contributions to the total deviation, each of equal magnitude a and equally likely to be positive or negative. If we repeat the measuring process many times, we shall obtain a set of values that will range from $X + ma$, for a reading in which all the fluctuations happened to be positive simultaneously, to $X - ma$, if the same happened in the negative direction. For such random summation of positive and negative quantities (as in the "random walk"), we can prove that the most probable sum is zero, meaning that the most common values of x are in the vicinity of X. The distribution curve, therefore, has a peak in the middle, is symmetrical, and declines smoothly to zero at $x = X + ma$ and $x = X - ma$. If this concept is taken to the limiting case in which an infinite number of infinitesimal deviations contribute to the total deviation, the curve has the form shown in Fig. 3–3. Treating the curve solely from the mathematical point of view for the moment, its equation can be written

$$y = Ce^{-h^2(x-X)^2} \tag{3–3}$$

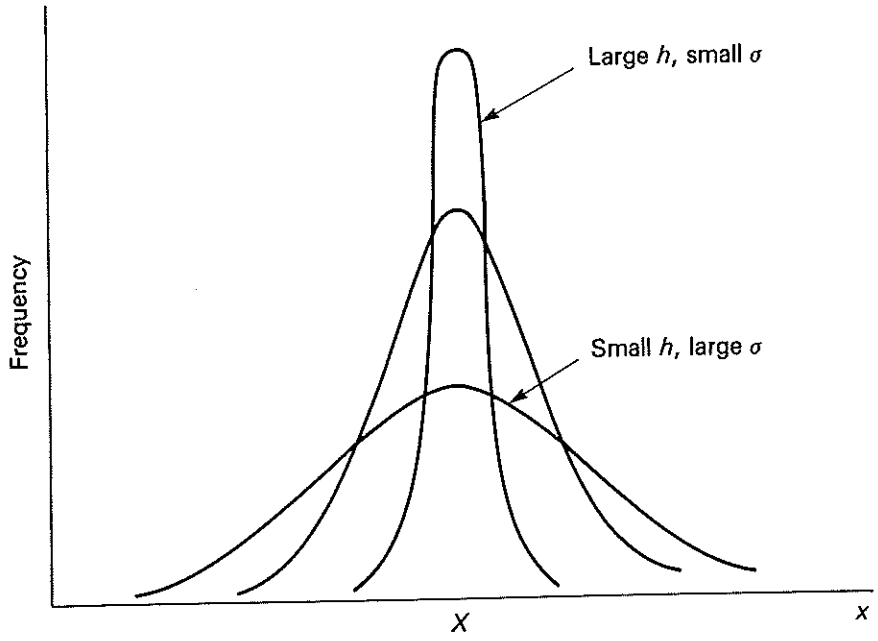

Figure 3-3 The Gaussian distribution curve.

Here the constant C is a measure of the height of the curve, since $y = C$ for $x = X$ at the center of the distribution. The curve is symmetrical about $x = X$ and approaches zero asymptotically. The quantity h obviously governs the width of the curve, since it is only a multiplier on the x scale. If h is large, the curve is narrow, and high in relation to its width; if small, the curve is low and broad. The quantity h clearly must be connected with the standard deviation, σ, of the distribution, and it can be shown that the relationship is

$$\sigma = \frac{1}{\sqrt{2}\,h} \tag{3–4}$$

(We shall use Latin letters, e.g., S for standard deviation, for quantities associated with finite sets of actual observations, and Greek letters, such as σ, when referring to defined distributions or, as described in Sec. 3–7, to a "universe" of observations.)

Now that we have a definite equation for the distribution, all the original ambiguity about interpreting the standard deviation in terms of probability disappears, and we have definite, unique, and permanent values. For example, the area enclosed within the interval $X \pm \sigma$ for a Gaussian distribution is 68% and within the interval $X \pm 2\sigma$ it is 95%, and this is so for *all* Gaussian distributions. The relation between the σ values and areas on the distribution curve is shown in Fig. 3–4 by the lines drawn vertically at intervals of 1σ and 2σ from the central value. It is very comforting to have such definite numbers, be-

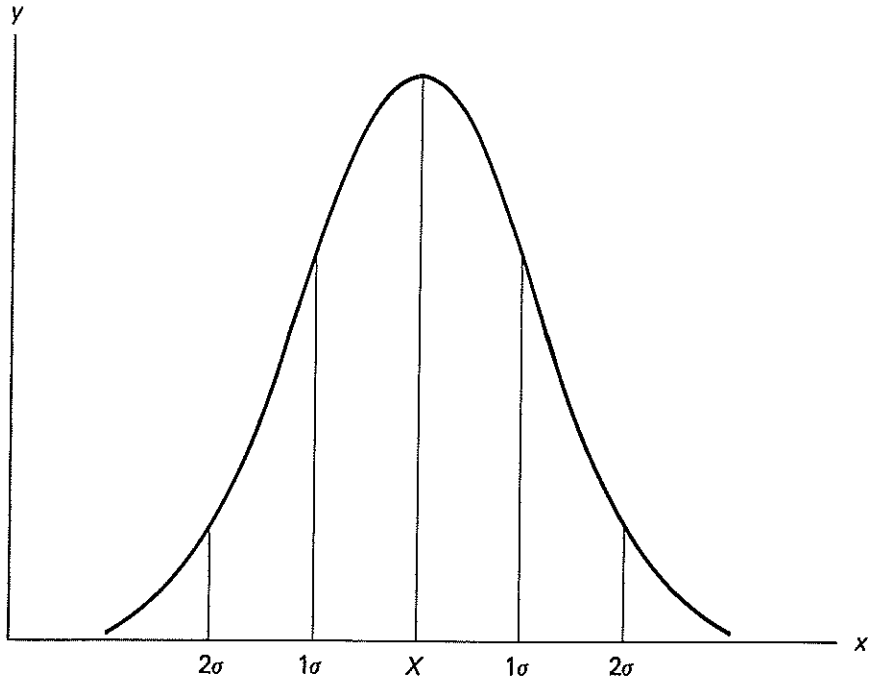

Figure 3-4 The relationship of 1σ and 2σ limits to the Gaussian distribution.

cause we can say definitely that any particular value in a Gaussian set has a 68% chance of falling within the interval $X \pm \sigma$ and a 95% chance of falling within $X \pm 2\sigma$. A more extensive account of the mathematical properties of the Gaussian distribution will be found in Appendix 1.

3–7 RELATION BETWEEN GAUSSIAN DISTRIBUTIONS AND REAL OBSERVATIONS

The results given in the preceding section provide useful, precise methods for interpreting means and standard deviations, but a problem arises when we start applying such thoughts to real measurements. Numbers like 68% and 95% refer to a theoretical construct, the Gaussian distribution, and all we have is one, or at most a few, real measurements of our desired quantity. We have, at first, no way of knowing which Gaussian distribution, with attached values of X and σ, is appropriate to our observations. So what are we to do? The answer lies in a concept which provides a bridge between the world of theoretical constructs and the world of real measurements. We invent, for a piece of apparatus or a measuring process, the concept of the infinite set of measurements which *could* be made with it. Of course, for rather obvious reasons, this infinite set of measurements will never be made, but the concept enables us to interpret our real measurements. The construct is called the "universe" or

"population" for that particular measurement. Once we have made, say, 100 measurements with a particular apparatus, we have a tendency to feel that nothing exists but our 100 values. We must now invert our thinking and view our set of measurements as a "sample" of the infinitely large universe or population of measurements that could be made. The universe, however, is permanently inaccessible to us; we shall never know the universe distribution or its mean or its standard deviation. Our task will be to construct inferences about these quantities from the definitely known properties of our sample.

We shall do this on the basis of some assumptions. First, we shall assume that the universe distribution is Gaussian, and we shall call the universe mean X and the universe standard deviation σ. This assumption enables us to make statements such as: if we make just one measurement with our equipment, that one measurement has a 68% chance of falling within $X \pm \sigma$ and a 95% chance of falling within $X \pm 2\sigma$. This seems like an encouragingly exact and explicit statement, but it suffers from an overwhelming defect; we do not (and never shall) know the values of X and σ. In other words, having made only one observation of a quantity that is subject to random fluctuation, we have gained practically nothing. We can say only that our value has a 68% chance of falling within something of somewhere, which is not too helpful. Our only hope lies in obtaining some information, even if uncertain, about the universe distribution. As we have already mentioned, we are never going to be able to determine the universe distribution exactly, because that would require an infinite number of readings. We can only hope that, if we repeat our measuring process to obtain a sample from the universe, that sample will enable us to make some estimate of the universe parameters.

Since we are making the basic assumption that the universe distribution is a mathematical, defined function (whether Gaussian or some other, equally well-defined distribution), we can evaluate mathematically the properties of samples with respect to those of the universe of single observations. We shall simply state these properties of samples without proof. The reader who is curious about the mathematical derivation of these results is encouraged to turn to the standard texts on statistics, in which sections dealing with sampling theory will be found.

The properties of samples become clear if we consider the concept of repeated sampling. Consider that, with a certain piece of apparatus, we make 100 observations. This will be our first sample; let us calculate its mean and standard deviation and record them. Now let us make another set of 100 observations and record for it the mean and standard deviation. Let us continue such repetition until we have an infinite number of samples, each with its own mean

and standard deviation, and let us then plot the distribution curves of the sample means and of the sample standard deviations. Of course we shall never carry out a process like this with actual observations but, knowing the mathematical function for our original universe of single readings, we can simulate such repeated sampling mathematically, and so derive the properties of the samples in comparison with those of the original universe of single readings. The results of such calculations of the distribution of sample means and sample standard deviations are shown in Fig. 3–5 and Fig. 3–6 and they will be described in the following sections.

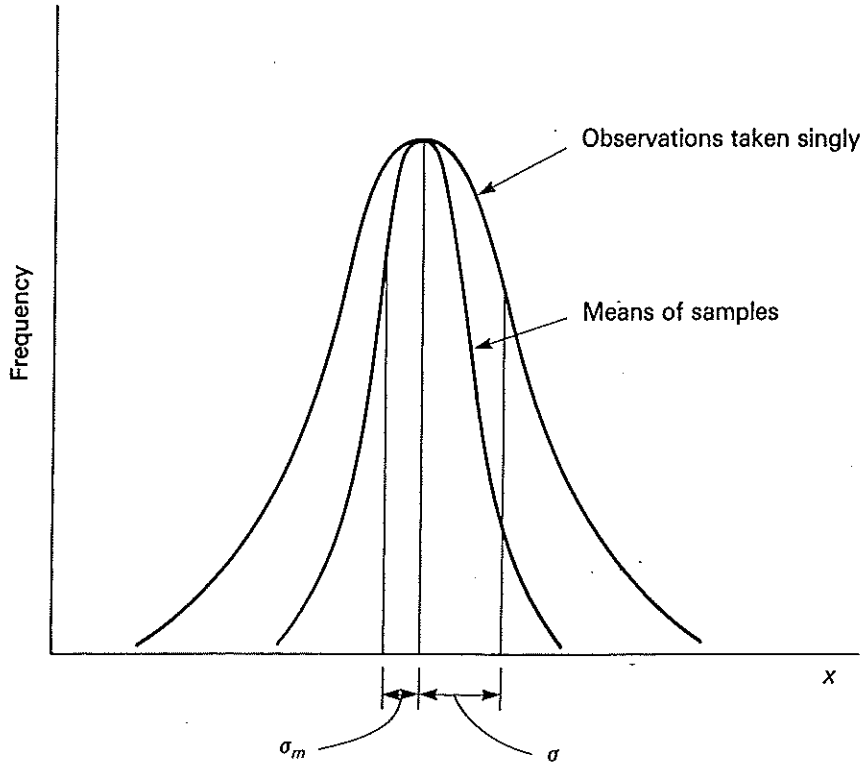

Figure 3-5 Distribution curve of single observations and sample means. (Note that the vertical scale for the two curves is not the same. They have been plotted with a common peak value solely for purposes of illustration.)

3–8 SAMPLE MEANS AND STANDARD DEVIATION OF THE MEAN

If the universe destribution of single readings is Gaussian, the theory of sampling shows that the distribution of sample means is also Gaussian. In addition, the distribution of sample means has two other very important properties. First,

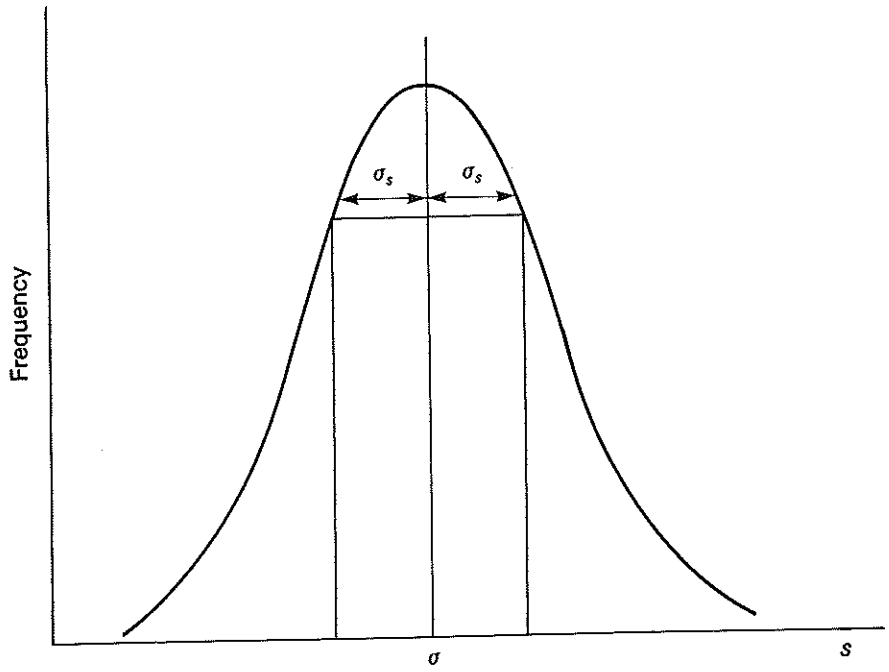

Figure 3-6 The distribution of sample standard deviations.

it is centered on X, the center of the original distribution of single readings; second, it is narrower than the original distribution. This narrowness is very significant, because it demonstrates immediately the improvement in precision that comes from samples as opposed to single readings; the means of samples cluster more closely around the universe mean than do single readings. The reduced scatter of sample means is represented by a very important quantity, the standard deviation of the distribution of sample means. This quantity is called the **standard deviation of the mean** and its value is

$$\sigma_m = \frac{\sigma}{\sqrt{N}} \tag{3–5}$$

where N is the number of readings in the sample. Thus, a particular sample mean has a 68% chance of falling within the interval $X \pm \sigma_m$ and a 95% chance for the interval $X \pm 2\sigma_m$. These intervals are smaller than the corresponding intervals for single readings, and they supply a numerical measure of the improved precision that is available from sampling.

Note that the statement about sample means, although precise, still does not help us very much, because it still involves the unknown quantities X and σ. The resolution of this difficulty and the significance of the standard deviation of the mean will become clear very soon. In the meantime let us turn our attention briefly to the other important property of samples, the distribution of sample standard deviations.

3–9 SAMPLE STANDARD DEVIATION

The sample standard deviations also fall on a Gaussian distribution, the center of which is the universe standard deviation, σ. The distribution is illustrated in Fig. 3–6. As will become clear in a moment, however, the variance of the sample standard deviations will not concern us as much as the variance of sample means, and we shall postpone to Sec. 3–11 further discussion of the variance of sample standard deviations.

3–10 APPLICATION OF SAMPLING THEORY TO REAL MEASUREMENTS

The sample properties which we have just presented are very interesting, but how do they help us when we do not have access to the actual distributions, either for sample means or sample standard deviations? We have our lone sample with *its* mean and standard deviation, and no idea how they relate to the universe values. Our problem, therefore, is to find a connection between the theoretical results and the sample properties that allows us to infer the universe properties from the sample values. We cannot expect, obviously, to obtain exact information. In addition, we must make one basic, obviously imprecise, assumption. We assume that our single value, the sample standard deviation, provides us with a value for the universe standard deviation. In fact it can be proved that the "best estimate" of the universe standard deviation is given by the quantity

$$S = \sqrt{\frac{\Sigma\,(\bar{x} - x_i)^2}{N - 1}} \tag{3–6}$$

This quantity is only slightly different from our original value for the standard deviation of a set of observations. The N in the denominator of the original expression has been replaced by $N - 1$, and the difference between the two quantities, obviously, is significant only for small values of N. In the future, when we talk about a sample standard deviation, we shall assume that we are using the equation in the new form and that we are really talking about the "best estimate" of the universe value σ.

Accepting our sample standard deviation as the best estimate of σ, we are now able to make a definite statement about our single sample. We can rephrase Eq. (3–5) and define

$$S_m = \frac{S}{\sqrt{N}} \tag{3–7}$$

as our standard deviation of the mean, now a known quantity obtained from our real sample. We can now say: our sample mean $\bar{x}$ has a 68% chance of

falling within $X \pm S_m$ and a 95% chance of falling within $X \pm 2S_m$. This is a statement which is close to what we want, but it is not yet completely satisfactory. It tells us something about a quantity that we know, $\bar{x}$, in terms of a quantity that we do not know, X. We really want the statement to be the other way around; we want to be able to make an assertion about our unknown, X, in terms of a quantity, $\bar{x}$, of which we do know the value. Fortunately, it is possible to prove that the above statement about probabilities can be inverted to yield our desired result. We obtain thereby the statement toward which we have been working ever since we started our discussion of the statistics of fluctuating quantities. Our final statement is: there is a 68% chance that the universe mean, X, falls within the interval $\bar{x} \pm S_m$ and a 95% chance that it falls within the interval $\bar{x} \pm 2S_m$. This is now, finally, a statement about the unknown quantity, X, in terms of wholly known quantities, $\bar{x}$ and S_m. Along our scale of x values we now have a real and known interval between $\bar{x} - S_m$ and $\bar{x} + S_m$, and we know that there is a 68% chance that our desired quantity X lies within this interval.

This statement provides us with the answer we have been seeking and brings us as close as we can come to exact information about the unperturbed value of our measured quantity. It is worth becoming familiar with the arguments that have been given in the preceding sections; there is more to measurement than simply making a few measurements and "taking the average" just because it seems to be the right thing to do.

3–11 EFFECT OF SAMPLE SIZE

In any sampling process, clearly, the larger the sample, the more precise will be our final statements. Even though the precision of a mean value increases only as the square root of the number of observations in the sample [Eq. (3–5)], it does increase, and larger samples have more precise means. There may, however, be limitations of time or opportunity, and we cannot always obtain samples of the size we would like. Usually a compromise must be sought between the conflicting demands of precision and time, and good experiment design will incorporate this compromise into the preliminary planning. Nevertheless, it may occasionally be necessary to content ourselves with small samples. In this undesirable eventuality we should be aware of the magnitude of the resulting loss of precision. There is, first, the influence on the value of the standard deviation of the mean; the smaller N is, the larger will be the value of S_m and the longer the interval on the x scale that has the 68% chance of containing the universe value X. Second, we must, for small samples, place declining faith in our use of the sample standard deviation S as the best estimate of the universe value σ. To illustrate this, recall the distribution curve for sample standard deviations that was shown in Fig. 3–6. It is worth asking:

given the existence of this distribution, how good is our "best estimate" of the universe standard deviation, and how does it vary with sample size? The answer must be based on the width of the distribution of sample standard deviations, and so we must calculate the standard deviation of this distribution. It is called the **standard deviation of the standard deviation.** (This process could obviously go on indefinitely but, thankfully, we shall stop at this stage.) The value of the standard deviation of the standard deviation, calculated mathematically from the equation of the Gaussian distribution, is

$$\sigma_S = \frac{\sigma}{\sqrt{2(N-1}} \tag{3–7}$$

The breadth of the distribution of sample standard deviations is thus related to its central value σ by the numerical factor $1/\sqrt{2(N-1)}$. As one might expect, therefore, the accuracy of our sample standard deviation as the best estimate of the universe value is dependent on the sample size. For example, with a sample size of 10, Eq. (3–8) shows that our S value from the sample has a 68% chance of falling within a range of $\pm\sigma/\sqrt{18}$, approximately $\pm\sigma/4$, about the universe value σ. Correspondingly, the interval that has a 95% chance of containing our sample mean is as wide as $\sigma/2$ about the universe value σ. This does not represent high precision of measurement. We have, therefore, confirmation of the warning given earlier: statistical exercises with small samples should be undertaken only when no alternative exists. In order to provide an overall feeling for the reliability of σ estimates from samples of differing size, Table 3–1 contains some typical values of $\sqrt{2(N-1)}$ for various values of N.

TABLE 3–1 Accuracy of σ Estimates from Samples of Varying Size

68% Confidence		95% Confidence	
N	$\sqrt{2(N-1)}$	N	$\sqrt{2(N-1)}$
2	1.4	2	0.7
3	2.0	3	1.0
4	2.4	4	1.2
5	2.8	5	1.4
6	3.1	6	1.6
7	3.4	7	1.7
8	3.7	8	1.8
9	4.0	9	2.0
10	4.2	10	2.1
15	5.2	15	2.6
20	6.1	20	3.2
50	9.8	50	4.9
100	14.1	100	7.0

These values are illustrated in Fig. 3–7 for $N = 3$, $N = 10$, and $N = 100$. The $\pm 1\sigma/S$ limits are marked on these curves, showing, for various sample sizes, the intervals within which there is a 68% probability that our single sample standard deviation lies. For values of N less than about 10, it is clear that the intervals for 68% or 95% probability become so large in comparison with the central value that it is almost pointless to attempt an estimate of σ. It is, therefore, rarely worth attempting any kind of statistical analysis with samples containing fewer than about 10 observations. In any case, when reporting the outcome of statistical work, it is essential to quote the sample size. If we intend our values for the mean and standard deviation of the mean to be interpreted in accordance with the 68% and 95% prescription, we must give our reader the opportunity to judge the accuracy of our estimates.

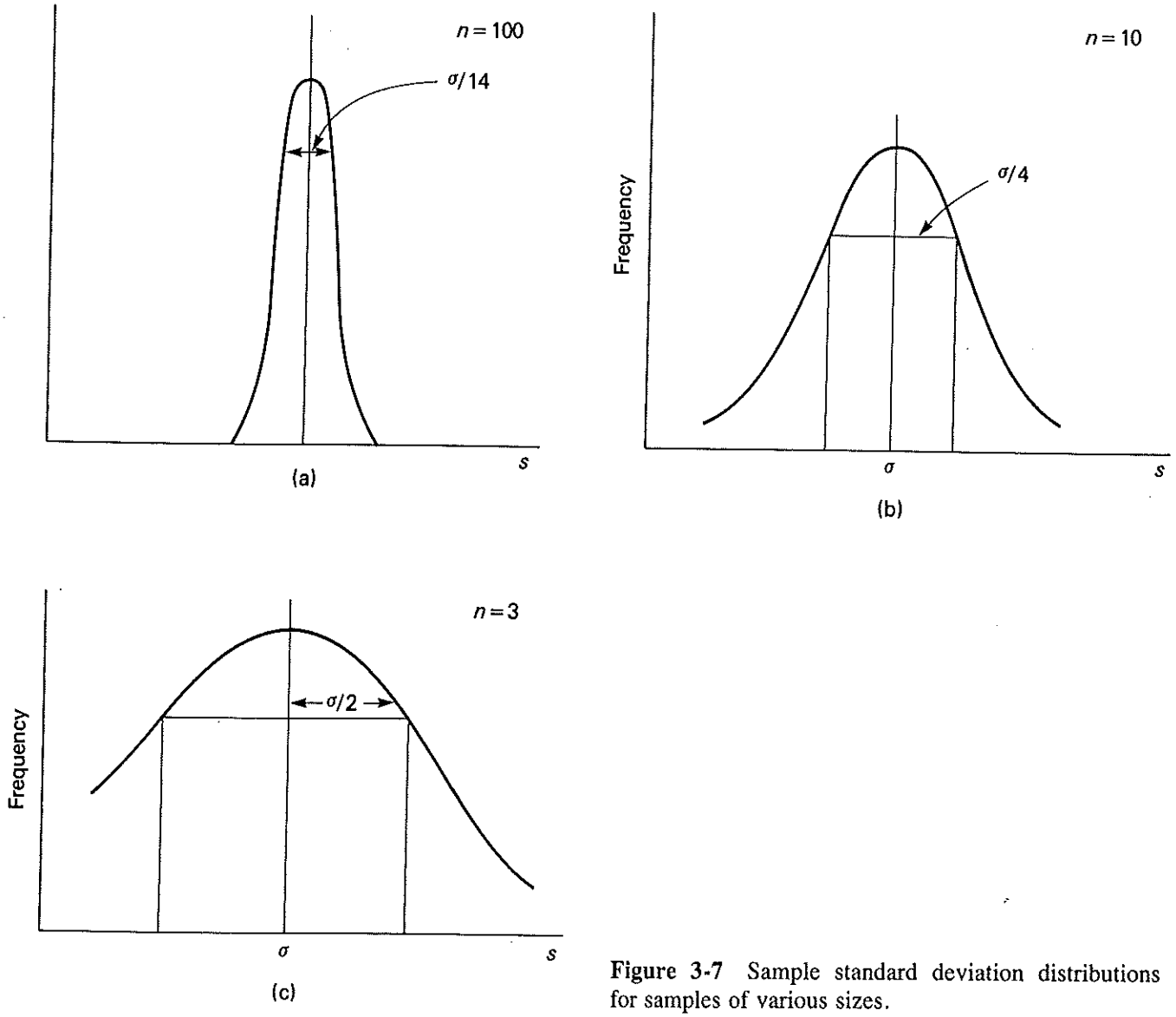

Figure 3-7 Sample standard deviation distributions for samples of various sizes.

3–12 STANDARD DEVIATION OF COMPUTED VALUES

In Chap. 2 we considered the uncertainty of computed values z, and we assumed that the uncertainty of the basic measurements constituted intervals within which we were almost certain that the values lay. We calculated the maximum range of variability of the computed answer on the pessimistic assumption that the errors in the various measured values combined, in a "worst-case" fashion, to drive the computed answer as far away from the central value as it can go. We have already suggested that this represents an unrealistically pessimistic approach and that a more useful quantity would be a "probable" value for δz based on the various probabilities associated with deviation of the basic quantities x, y, etc. from their central values. The limits given by this quantity will, naturally, be smaller than $\pm\delta z$, but we hope to find actual, numerical significance for them. Such statistical validity will be available only if the uncertainties in x and y have statistical significance, and we shall assume in the following calculations that the measurements of x and y have been sufficiently numerous to justify a calculation of the standard deviaions S_x and S_y. We hope now to calculate a value for S_z that will have the same significance for z values as S_x and S_y had for x and y.

We must, however, first inquire what we mean by S_z. We assume that the measurement has taken the form of pairs of observations x, y that were obtained by repetition of the observing process under identical conditions (for example, the current through and the potential across a resistor, measured for the purpose of calculating the resistance R). Each pair of observations will provide a value of z, and, if repetition yielded N pairs, we shall have a set of N values of z that are distributed in accordance with the fluctuations in the basic measurements. The quantity we require, S_z, is the standard deviation of this set of z values. These individual values of z may never be calculated individually, because a simpler mode of calculation exists. We can calculate the means $\bar{x}$ and $\bar{y}$, of the sets of x and y values and obtain $\bar{z}$ directly using the assumption (valid if S_x, S_y, and S_z are small compared, respectively, with $\bar{x}$, $\bar{y}$, and $\bar{z}$) that

$$\bar{z} = f(\bar{x}, \bar{y})$$

Nevertheless, that distribution of z values provides the significance of the S_z that we are about to calculate.

If we assume that the universes of the x, y, and z values have a Gaussian distribution, the quantity σ_z (of which we are about to calculate the best estimate in terms of various S values) will have the usual significance; i.e., any z value will have a 68% chance of falling within $\pm\sigma_z$ of the central value. As be-

fore, let

$$z = f(x, y)$$

and consider perturbations δx and δy which lead to a perturbation δz in the computed z value. The value of δz will be given by

$$\delta z = \frac{\delta z}{\delta x}\,\delta x + \frac{\delta z}{\delta y}\,\delta y$$

This perturbation can be used to calculate a standard deviation for the N different z values, since

$$S_z = \sqrt{\frac{\Sigma\,(\delta z)^2}{N}}$$

Thus

$$\begin{aligned} S_z^2 &= \frac{1}{N}\sum\left(\frac{\partial z}{\partial x}\,\delta x + \frac{\partial z}{\partial y}\,\delta y\right)^2 \\ &= \frac{1}{N}\sum\left(\left(\frac{\partial z}{\partial x}\right)^2(\delta x)^2 + \left(\frac{\partial z}{\partial y}\right)^2(\delta y)^2 + 2\,\frac{\partial z}{\partial x}\,\frac{\partial z}{\partial y}\,\delta x\,\delta y\right) \\ &= \left(\frac{\partial z}{\partial x}\right)^2\frac{1}{N}\sum(\delta x)^2 + \left(\frac{\partial z}{\partial y}\right)^2\frac{1}{N}\sum(\delta y)^2 + \frac{2}{N}\,\frac{\partial z}{\partial x}\,\frac{\partial z}{\partial y}\,\delta x\,\delta y \end{aligned}$$

But

$$\frac{1}{N}\sum(\delta x)^2 = S_x^2 \quad \text{and} \quad \frac{1}{N}\sum(\delta y)^2 = S_y^2$$

and, since δx and δy may be considered for the present purpose to be independent perturbations,

$$\sum(\delta x\,\delta y) = 0$$

Thus, finally,

$$S_z = \sqrt{\left(\frac{\partial z}{\partial x}\right)^2 S_x^2 + \left(\frac{\partial z}{\partial y}\right)^2 S_y^2} \tag{3–8}$$

If z is a function of more than two variables, the equation is extended by adding similar terms. Thus, if the components of a calculation have standard deviations with some degree of reliability, a value can be found for the probable uncertainty of the answer, where "probable" has real numerical significance.

The calculation has been performed in terms of the variance or standard deviation of the x and y distributions. In actual practice, however, we do not use the sample variance directly; we must calculate the best estimates of σ_x, σ_y, etc., and, in accordance with Eq. (3–6), we would use the modified value for standard deviation with denominator $N - 1$ instead of N. The final result would then be a best estimate for σ_z. The standard deviation of the mean for z can then be calculated by direct use of Eq. (3–5) and will give the limits that have a 68% chance of containing the unperturbed value.

Note that most experiments are not carried out in accordance with the restricted assumptions of the above development. If, for example, we are studying the flow rate of water through a pipe, we would measure the flow rate, pipe radius, and pipe length independently and would choose the number of readings in each sample on the basis of the intrinsic precision of the measurement. We cannot, therefore, use Eq. (3–8) directly, since the various S's are not compatible. The solution is to calculate the standard deviation of the mean for each of the elementary quantities first. If these are used in Eq. (3–8), the result of the calculation will be immediately a standard deviation of the mean for z.

3–13 STANDARD DEVIATION OF COMPUTED VALUES: SPECIAL CASES

Let us now apply Eq. (3–8) to a few common examples. In all the following cases the various S's are all assumed to be best estimates of the appropriate universe value σ.

(a) Sum of Two Variables

$$z = x + y$$

Here

$$\frac{\partial z}{\partial x} = 1, \qquad \frac{\partial z}{\partial y} = 1$$

and

$$S_z = \sqrt{S_x^2 + S_y^2}$$

Note that this result provides justification for Eq. (3–5). The mean value for the sample, $\Sigma\,(x_i/N)$, is just such a function as $z = x + y$, where x and y hap-

pen to be independent measurements of the same quantity. Thus, if

$$z = \frac{1}{N}(x_1 + x_2 + x_3 + \cdots)$$

$$\frac{\partial z}{\partial x_1} = 1/N \qquad \frac{\partial z}{\partial x_2} = 1/N \qquad \text{etc.}$$

and

$$S_z = \sqrt{\left(\frac{1}{N}\right)^2 S_x^2 + \left(\frac{1}{N}\right)^2 S_x^2 + \cdots}$$
$$= \sqrt{NS_x^2/N^2} = S_x/\sqrt{N}$$

(b) Difference of Two Variables

$$z = x - y$$

Here

$$\frac{\partial z}{\partial x} = 1, \qquad \frac{\partial z}{\partial y} = -1$$

but, again,

$$S_z = \sqrt{S_x^2 + S_y^2}$$

Recalling Sec. 2–8(b), we note that the earlier discussion of measured differences is still valid.

(c) Product of Two Variables

$$z = xy$$

Here

$$\frac{\partial z}{\partial x} = y, \qquad \frac{\partial z}{\partial y} = x$$

and

$$S_z = \sqrt{y^2 S_x^2 + x^2 S_y^2}$$

The specific value of S_z at the particular values, x_0 and y_0, of x and y can be obtained by substituting x_0 and y_0 in the equations. As was the case for uncertainty in products, the equation is more clearly expressed in terms of relative values of S. We obtain

$$\frac{S_z}{z} = \sqrt{\frac{S_x^2}{x^2} + \frac{S_y^2}{y^2}}$$

(d) Variables Raised to Powers

Here

$$z = x^a$$

$$\frac{\partial z}{\partial x} = ax^{a-1}$$

and

$$S_z = \sqrt{a^2 x^{2(a-1)} S_x^2}$$

Again this is more instructive when expressed in terms of relative values:

$$\frac{S_z}{z} = \sqrt{\frac{a^2 S_x^2}{x^2}}$$

$$= a\frac{S_x}{x}$$

(e) The General Case of Powers and Products

$$z = x^a y^b$$

The results of the two preceding sections can obviously be extended to give the result

$$\frac{S_z}{z} = \sqrt{\left(\frac{aS_x}{x}\right)^2 + \left(\frac{bS_y}{y}\right)^2}$$

In contrast to the case of combined uncertainty, negative powers in the original function need not be given special consideration; in the equation for S_z powers occur in squared form and automatically make a positive contribution.

If a function other than those listed above is encountered, the use of Eq. (3–8) will yield the desired result. Incidentally, we may note that, for a function of a single variable, Eq. (3–8) reduces to the same form as for uncertainties, Eq. (2–1). This correspondence is predictable for a situation in which we do not have the probability-based interplay between two or more variables.

Finally, although we listed in Sec. 2–5 to Sec. 2–9 a number of different approaches to the calculation of outer limits for uncertainties, the standard deviation of z is a uniquely defined quantity, and there is no alternative to the use of Eq. (3–8).

3–14 COMBINATION OF DIFFERENT TYPES OF UNCERTAINTY

Unfortunately for the mathematical elegance of the development, we frequently require the uncertainty in a computed result which contains quantities having different types of uncertainty. We may require the uncertainty in

$$z = f(x, y)$$

where, for example, x is a quantity to which have been assigned outer limits, $\pm \delta x$, within which we are "almost certain" that the actual value lies, while y is a quantity whose uncertainty is statistical in nature, a sample standard deviation, S_y, perhaps, or a standard deviation of the mean, $S_y/\sqrt{N}$. We require an uncertainty for z. Our initial difficulty is even to define the uncertainty in z. We are trying to combine two quantities which have, in effect, completely different distibution curves. One is the standard Gaussian function; the other is a rectangle, bounded by the values $x_0 + \delta x$ and $x_0 - \delta x$ and flat on top, because the actual value of x is equally likely to be anywhere within the interval $x_0 \pm \delta x$. Any general method of solving this problem is likely to be far too complex for general use, but a simple approximation is available using the following procedure.

In the calculation for z we use the sample mean, $\bar{y}$, for the y value, implying that the universe mean has an approximately $\frac{2}{3}$ chance of falling within the interval, $\bar{y} \pm S_y/\sqrt{N}$. Let us, therefore, calculate limits for x that also have a $\frac{2}{3}$ probability of enclosing the actual value. Since the probability distribution for x is rectangular, $\frac{2}{3}$ of the area under the distribution curve is enclosed by limits that are separated by a distance equal to $\frac{2}{3}$ of the total range of possibility, i.e., $\frac{2}{3}$ of 2 δx. The total width of the region for $\frac{2}{3}$ probability is, therefore, $\frac{4}{3}$ δx and the uncertainty limits are $\pm \frac{2}{3}$ δx.

The quantity $\frac{2}{3}$ δx is, therefore, compatible with $S_y/\sqrt{N}$, since both refer to $\frac{2}{3}$ probability. Equation (3–9) can now be used, inserting $\frac{2}{3}$ δx for the value of the standard deviation of the mean for x and $S_y/\sqrt{N}$ for the y function. This will yield a value for uncertainty in z which can be interpreted in accordance with the $\frac{2}{3}$ prescription. Note, however, that the limits for 95% probability are not simply twice as wide as those for $\frac{2}{3}$ probability; they would have to be calculated separately using the above method.

3–15 REJECTION OF READINGS

One last, practical property of distribution curves concerns outlying values. There is always the possibility of making an actual mistake, perhaps in misreading a scale or in accidentally moving an instrument between setting and

reading. We encounter the temptation, therefore, to assign some such cause to a single reading that is well separated from an otherwise compact group of values. This is, however, a dangerous temptation, since the Gaussian curve does permit values remote from the central part of the curve. Furthermore, once we admit the possibility of pruning the observations, it can become very difficult to know where to stop. We are dependent, therefore, on the judgment of the experimenter. This is not unreasonable, since the experimenter knows more about the measurement than anyone else, but criteria for making the choices can be helpful. Many empirical "rules" for rejection of observations have been formulated, but they must be used with discretion. It would be foolish to use a rule to reject one reading which was just outside the limit set by the rule if there are other readings just inside it. There is also the possibility that extra information relating to the isolated reading was noted at the time it was made, and this can help us decide in favor of retention or rejection.

The guidance we desire in making such decisions can be found in the properties of the Gaussian distribution. In a Gaussian distribution the probability of obtaining readings outside the 2σ limits is 5% (as we have seen before), outside 3σ limits it is $\frac{1}{3}\%$, and outside 4σ limits the chance is no more than 6×10^{-5}. The decision to reject is still the responsibility of the experimenter, of course, but we can say, in general terms, that readings falling outside 3σ limits are likely to be mistakes and candidates for rejection. However, a problem can arise because of our lack of information about the universe of readings and its parameters X and σ. The better our knowledge of σ, the more confident we can be that any far-out and isolated reading arises from a genuinely extraneous cause such as personal error, malfunction of apparatus, etc. Thus, if we make 50 observations that cluster within 1% of the central value and then obtain one reading that lies at a separation of 10%, we can be fairly safe in suggesting that this last reading did not belong to the same universe as the preceding 50. The basic requirement, before any rejection is justified, is confidence in the main distribution of readings. Clearly, there is no justification for taking two readings and then rejecting a third measurement on the basis of a 3σ criterion. Unless the case for rejection is completely convincing, the best course is to retain all readings, whether we like them or not.

It is wise also to remember that many of the greatest discoveries in physics had their origin in outlying measurements.

PROBLEMS

The following observations of angles (in minutes of arc) were made while measuring the thickness of a liquid helium film. Assume that the observations show random un-

certainty, that they are a sample from a Gaussian universe, and use them in Problems 1–14.

34	35	45	40	46
38	47	36	38	34
33	36	43	43	37
38	32	38	40	33
38	40	48	39	32
36	40	40	36	34

1. Draw the histogram of the observations.
2. Identify the mode and the median.
3. Calculate the mean.
4. Calculate the best estimate of the universe standard deviation.
5. Calculate the standard deviation of the mean.
6. Calculate the standard deviation of the standard deviation.
7. **(a)** Within which limits does a single reading have a 68% chance of falling? **(b)** Which limits give a 95% chance?
8. Within which limits does the mean have **(a)** a 68% chance, and **(b)** a 95% chance of falling?
9. Within which limits does the sample standard deviation stand **(a)** a 68% chance and **(b)** a 95% chance of falling?
10. Calculate a value for the constant h in the equation for the Gaussian distribution.
11. If a single reading of 55 had been obtained in the set, would you have decided in favor of accepting it or rejecting it?
12. Take two randomly chosen samples of five observations each from the main set of readings. Calculate their sample means and standard deviations to see how they compare with each other and with the more precise values obtained from the big sample.
13. If the experiment requires that the standard deviation of the mean should not exceed 1% of the mean value, how many readings are required?
14. If the standard deviation of the universe distribution must be known within 5%, how many readings are required?
15. Repeated measurements of the diameter of a wire of circular cross section gave a mean of 0.62 mm with a sample standard deviation of 0.04 mm. What is the standard deviation for the calculated value of the cross-sectional area?
16. The wavelength of the two yellow lines in the sodium spectrum are measured to be 589.11×10^{-9} m and 589.68×10^{-9} m, each with a standard deviation of 0.15×10^{-9} m. What is the standard deviation for the calculated difference in wavelength between the two lines?
17. A simple pendulum is used to measure g using $T = 2\pi\sqrt{l/g}$. Twenty measurements of T give a mean of 1.82 sec and a sample standard deviation of 0.06 sec. Ten measurements of l give a mean of 0.823 m and a sample standard deviation of 0.014 m. What is the standard deviation of the mean for the calculated value of g?